MAT 2410: Optimisation

Robert Guénette

Département de Mathématiques et de Statistique

Chapitre 4
Optimisation différentiable avec contraintes

Références

Condition d'optimalité: cas convexe

Dualité

- Lagrangien et point de selle
- Fonction indicatrice
- Contraintes $Bx = c$
- Contraintes $Bx \leq c$
- Cône des directions admissibles

Méthodes numériques

- contraintes d'égalité

Références:

- Notes de cours: chapitre 4, sections 4.1 et 4.2.
- Livre de M. Delfour: chapitre 4.

Problèmes de minimisation avec contraintes

Soit $f : U \longrightarrow \mathbb{R}$ une fonction numérique définie sur un ensemble de $\mathbb{R}^n$. Il s'agit de déterminer le point $a \in U$ qui vérifie

$$f(a) = \min_{x \in U} f(x)$$

En général U est définie par l'intersection d'une famille de contraintes de la forme:

- $g(x) \leq 0$,
- $h(x) = 0$,
- $Bx \leq c$ où B est une matrice et c un vecteur,
- $Cx = d$ où C est une matrice et d un vecteur,
- $lb \leq x \leq ub$ où les lb et ub sont des bornes sur la solution x, i.e. $lb_i \leq x_i \leq ub_i \quad \forall i$

Problèmes avec contraintes (suite)

Le type de problème que nous allons considérer par la suite, s'écrira de manière plus compact

$$\begin{array}{cc} \min & f(x) \\ \boldsymbol{g}(x) \leq 0 & \\ \boldsymbol{h}(x) = 0 & \end{array}$$

où $\boldsymbol{g} = (g_1, g_2, \ldots, g_m)$ sont les contraintes d'inégalité (en incluant les bornes sur x) et $\boldsymbol{h} = (h_1, h_2, \ldots, h_p)$ les contraintes d'égalité. Cette écriture inclut aussi les contraintes linéaires.

Remarques:

- $U = \{x \mid \boldsymbol{g}(x) \leq 0 \;\text{ et }\; \boldsymbol{h}(x) = 0\}$ n'est pas convexe en général.
- Si les g_i sont convexes et les h_j linéaires, alors U est convexe.

Condition d'optimalité du premier ordre: cas convexe

Théorème: Condition nécessaire d'optimalité du premier ordre

Soit $f : U \longrightarrow \mathbb{R}$ une fonction numérique de classe C^1 définie sur un ensemble **convexe** fermé $U \subset \mathbb{R}^n$. Si $a \in U$ est un point minimisant de f,

$$f(a) = \min_{x \in U} f(x),$$

alors

$$(\nabla f(a), x - a) \geq 0 \qquad \forall x \in U$$

Preuve: $a, x \in U \Longrightarrow a + t(x - a) \in U$ car U est convexe pour $0 \leq t \leq 1$. Par hypothèse, on a que

$$\begin{aligned} f(a + t(x-a)) - f(a) &\geq 0 \\ \lim_{t \to 0} \frac{f(a + t(x-a)) - f(a)}{t} &\geq 0 \\ (\nabla f(a), x - a)) &\geq 0 \end{aligned}$$

Condition d'optimalité (suite)

Remarques:

- Géométriquement parlant, la condition $(\nabla f(a), x-a) \geq 0$ s'interprète en disant que l'angle θ entre le vecteur $\nabla f(a)$ et $x-a$ est dans l'intervalle
$$-\pi/2 \leq \theta \leq \pi/2.$$
- Si $U = \mathbb{R}^n$, on retrouve la condition d'optimalité sans contrainte: $\nabla f(a) = 0$.
- Si le point minimisant a se trouve à l'intérieur de U, on a aussi $\nabla f(a) = 0$.
En effet il suffit de prendre $x = a \pm r\, y$ pour r suffisamment petit et $y \in \mathbb{R}^n$ quelconque. On introduit ce point dans la condition d'optimalité $(\nabla f(a), x-a) \geq 0$. Ceci implique
$$(\nabla f(a), \pm y) \geq 0 \Longrightarrow (\nabla f(a), y) = 0 \;\; \forall y \Longrightarrow \nabla f(a) = 0.$$

Condition d'optimalité cas convexe (suite)

Théorème: Condition nécessaire et suffisante d'optimalité du premier ordre

Soit $f : U \longrightarrow \mathbb{R}$ une fonction numérique convexe de classe C^1 définie sur un ensemble convexe fermé $U \subset \mathbb{R}^n$.
On a la caractérisation suivante:
$a \in U$ est un point minimisant de f si et seulement si

$$(\nabla f(a), x - a) \geq 0 \qquad \forall x \in U$$

Preuve: Il suffit de montrer $\Longleftarrow$
Par le critère de convexité du premier ordre, on a

$$f(x) \geq f(a) + (\nabla f(a), x - a) \qquad \forall x \in U.$$

Mais $(\nabla f(a), x - a) \geq 0$. On obtient

$$f(x) \geq f(a) + (\nabla f(a), x - a) \geq f(a) \qquad \forall x \in U.$$

Choix particuliers de U

- U est un sous-espace linéaire de $\mathbb{R}^n$. Dans ce cas, la condition d'optimalité s'écrit sous la forme

$$(\nabla f(a), y) = 0 \qquad \forall y \in U.$$

C'est-à-dire que $\nabla f(a) \in U^\perp$.

- U est un sous-espace affine de $\mathbb{R}^n$, i.e. $U = x_0 + W$ avec W un sous-espace linéaire. La condition d'optimalité s'écrit sous la forme

$$(\nabla f(a), w) = 0 \qquad \forall w \in W.$$

ou $\nabla f(a) \in W^\perp$ ou encore

$$(\nabla f(a), x - a) = 0 \qquad \forall x \in U.$$

Exemple

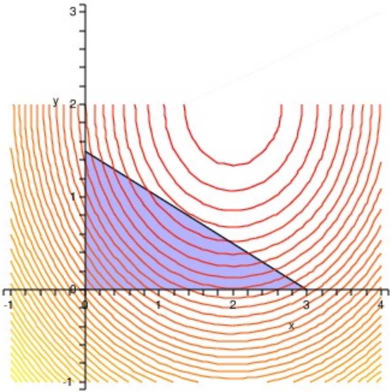
Considérons le problème de calculer la projection du point $(2,2)$ sur un polygone convexe. Le problème s'écrit sous la forme

$$\begin{array}{ll} \min & (x_1-2)^2+(x_2-2)^2 \\ x_1+2x_2 \leq 3 & \\ x_1 \geq 0, x_2 \geq 0 & \end{array}$$

Ce problème admet une solution unique. Selon le dessin, le gradient doit être dans la direction du vecteur normal à la droite $x_1+2x_2=3$.

$$(x_1-2, x_2-2) = t(-1,-2)$$

pour $t \geq 0$. On trouve
$t = 3/5 \Longrightarrow x_1 = 7/5, \ \ x_2 = 4/5.$



Orthogonalité

Soit E un sous-espace linéaire de $\mathbb{R}^n$. On définit le sous-espace orthogonal par

$$E^\perp = \{y \in \mathbb{R}^n \mid (y, x) = 0 \;\; \forall x \in E\}$$

Propriétés:

- $E^\perp$ est un sous-espace linéaire de $\mathbb{R}^n$,
- $dim\, E + dim\, E^\perp = n$,
- $E^{\perp\perp} = E$,
- $E \subset F \Longrightarrow F^\perp \subset E^\perp$,
- Si $B : \mathbb{R}^n \to \mathbb{R}^m$ est une application linéaire (matrice), on a

  $$(Ker\, B)^\perp = Im\, B^t$$

  où $Ker\, B = \{x \mid Bx = 0\}$.
  Rappel: le rang de B est égal à $rg(B) = dim\,(Im\, B)$ et on a $rg(B) = rg(B^t) = dim\,(Im\, B^t)$.

Cône dual

Notion de cône: $C \subset \mathbb{R}^n$ est un cône de sommet 0 si

$$x \in C, \ \ \lambda \geq 0 \Longrightarrow \lambda x \in C$$

Un cône convexe est caractérisé par

- $\forall x, y \in C \Longrightarrow x + y \in C$
- $\forall \lambda \geq 0, \ \forall x \in C \Longrightarrow \lambda x \in C$

Exemple important de cône convexe: $U = \{x \mid Bx \geq 0\}$ où B est une matrice de format $n \times m$.

Cône dual: soit $U \subset \mathbb{R}^n$ quelconque. On définit le cône dual par

$$U^* = \{y \mid (y, x) \geq 0 \ \ \forall x \in U\}$$

Propriétés:

- U^* est toujours un cône convexe fermé.
- $U \subset V \Longrightarrow V^* \subset U^*$
- $U^{**} = U$ si U est un cône convexe fermé.

Lagrangien et point de selle

Soient $A \subset \mathbb{R}^n$ et $B \subset \mathbb{R}^m$ deux ensembles non vide et

$$L : A \times B \to \mathbb{R}$$

une fonction appellée lagrangien.

Définition: Un point $(\bar{x}, \bar{y}) \in A \times B$ est un point de selle du lagrangien L si

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \qquad \forall x \in A, \quad \forall y \in B.$$

Exemple: Le point $(0, 0)$ est un point de selle de la fonction $L(x, y) = x^2 - y^2$ avec $A = B = \mathbb{R}$ car

$$L(0, y) = -y^2 \leq 0 = L(0, 0) \leq x^2 = L(x, 0) \qquad \forall x, y \in \mathbb{R}.$$

Théorème du min-max

Théorème: Si $(\bar{x}, \bar{y})$ est un point de selle de L, alors

$$\max_{y\in B}\min_{x\in A} L(x,y) = \min_{x\in A}\max_{y\in B} L(x,y) = L(\bar{x},\bar{y})$$

Autrement dit, on peut inverser l'ordre du min et du max.
Démonstration: Montrons en premier que

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y).$$

En effet, on a que

$$\min_{x\in A} L(x,y) \leq L(x,y) \leq \max_{y\in B} L(x,y)$$

Le membre de gauche est une fonction de y seulement. De même, pour le membre de droite qui est une fonction de x uniquement. Ceci implique

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \max_{y\in B} L(x,y), \qquad \forall x \in A$$

et, en prenant le minimum par rapport à x,

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y),$$

d'où le résultat.

Théorème du min-max (suite)

En deuxième, montrons que

$$\min_{x\in A}\max_{y\in B} L(x,y) \leq L(\bar{x},\bar{y}).$$

En effet, selon la définition du point de selle, on a

$$\max_{y\in B} L(x,y) = L(\bar{x},\bar{y}) \Longrightarrow L(\bar{x},\bar{y}) \geq \min_{x\in A}\max_{y\in B} L(x,y).$$

De même, on a

$$\min_{x\in A} L(x,y) = L(\bar{x},\bar{y}) \Longrightarrow L(\bar{x},\bar{y}) \leq \max_{y\in B}\min_{x\in A} L(x,y).$$

En recollant les morceaux, on obtient

$$L(\bar{x},\bar{y}) \leq \max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y) \leq L(\bar{x},\bar{y}),$$

d'où le résultat final.

Condition d'optimalité du lagrangien

Soient $U, V \subset \mathbb{R}^n$ deux sous-ensembles convexes et fermés. On fera les hypothèses suivantes:

- Pour y fixé, la fonction $L : x \to L(x, y)$ est convexe.
- Pour x fixé, la fonction $L : y \to L(x, y)$ est concave ($-L$ est convexe).

Au point-selle $(\bar{x}, \bar{y})$ du lagrangien L, la condition d'optimalité du minimum $L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in U$ s'écrit

$$(\frac{\partial L}{\partial x}(\bar{x}, \bar{y}), x - \bar{x}) \geq 0 \qquad \forall x \in U$$

De même, la condition d'optimalité du maximum $L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \quad \forall y \in V$ s'écrit

$$(\frac{\partial L}{\partial y}(\bar{x}, \bar{y}), y - \bar{y}) \leq 0 \qquad \forall y \in V$$

où $\frac{\partial L}{\partial x}$ est le gradient de la fonction $L(x, y)$ pour y fixé.

Fonction indicatrice

Soit $K \subset \mathbb{R}^n$ un sous-ensemble quelconque et non vide.
Definition: La fonction indicatrice de l'ensemble K est définie par

$$I_K(x) = \begin{cases} 0 & \text{si } x \in K, \\ \infty & \text{sinon.} \end{cases}$$

La fonction indicatrice sert à enlever les contraintes. Pour un problème de minimisation, on a évidemment la relation

$$\min_{x \in K} f(x) = \min_{x} f(x) + I_K(x)$$

Pour un problème de maximisation, on a plutôt

$$\max_{x \in K} f(x) = \max_{x} f(x) - I_K(x)$$

Contraintes de la forme: $Bx = c$

Appliquons la dualité lagrangienne au problème (primal)

$$\min_{Bx=c} \quad f(x)$$

où B est une matrice de format $m \times n$ et $c \in \mathbb{R}^m$. Ce problème a un sens seulement si $c \in Im\, B$ ou encore que le rang de B est maximal ($Im\, B = \mathbb{R}^m$).

Lemme: La fonction indicatrice de $K = \{x \in \mathbb{R}^n \mid Bx = c\}$ est donnée par

$$I_K(x) = \max_{\lambda \in \mathbb{R}^m} (\lambda, Bx - c)$$

Démonstration: en effet, si $Bx \neq c$, on peut prendre

$$\lambda = r(Bx - c) \Longrightarrow (\lambda, Bx - c) = r\|Bx - c\|^2 \to \infty \text{ si } r \to \infty$$

Le maximum sera ∞.
Au contraire, si $Bx = c$, on a que $(\lambda, Bx - c) = 0 \quad \forall \lambda \in \mathbb{R}^m$ de maximum 0.

Contraintes de la forme: $Bx = c$ (suite)

Par conséquent, on peut écrire

$$\min_{Bx=c} f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \in \mathbb{R}^m} (\lambda, Bx - c)$$

Ceci suggère la fonction lagrangienne

$$L(x, \lambda) = f(x) + (\lambda, Bx - c) \qquad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m.$$

ce qui transforme le problème primal en un problème de point-selle

$$\min_{Bx=c} f(x) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} L(x, \lambda)$$

La condition d'optimalité du lagrangien sera

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \Longleftrightarrow \nabla f(x) + B^t \lambda = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \Longleftrightarrow Bx = c \end{cases}$$

Point de vue des cônes duaux: $Bx = c$

La condition d'optimalité du problème $f(a) = \min_{Bx=c} f(x)$ s'écrit

$$(\nabla f(a), x - a) \geq 0 \qquad \forall x \in U = \{x \mid Bx = c\}.$$

Sachant que $c \in Im\, B$, on a que $U = x_0 + Ker\ B$ où $Bx_0 = c$. Ce qui implique que U est un sous-espace affine et dans ce cas, la condition d'optimalité correspond à

$$\nabla f(a) \in (Ker\ B)^{\perp} = Im\, B^t \Longrightarrow \qquad \exists \lambda \in \mathbb{R}^m \quad \nabla f(a) = -B^t \lambda.$$

Ceci fournit les conditions en disant que a et λ sont solutions du système

$$\begin{cases} \nabla f(x) + B^t \lambda = 0, \\ Bx = c. \end{cases}$$

Ce sont les mêmes conditions que celles obtenues par la méthode de Lagrange.

Contraintes de la forme: $Bx \leq c$

Appliquons la dualité lagrangienne au problème (primal)

$$\min_{Bx \leq c} \quad f(x)$$

où B est une matrice de format $m \times n$ et $c \in \mathbb{R}^m$.

Lemme: La fonction indicatrice de $K = \{x \in \mathbb{R}^n \mid Bx \leq c\}$ est donnée par

$$I_K(x) = \max_{\lambda \geq 0} (\lambda, Bx - c)$$

Démonstration: en effet, si $Bx > c$, on peut prendre

$$\lambda = r(Bx - c) \geq 0 \Longrightarrow (\lambda, Bx - c) = r\|Bx - c\|^2 \to \infty \text{ si } r \to \infty.$$

Le maximum sera ∞.
Au contraire, si $Bx \leq c$, on a que $(\lambda, Bx - c) \leq 0 \quad \forall \lambda \geq 0$ car $Bx - c \leq 0$. Donc le maximum sera 0.

Contraintes de la forme: $Bx \leq c$ (suite)

Par conséquent, on peut écrire

$$\min_{Bx \leq c} \; f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \geq 0} (\lambda, Bx - c)$$

Ceci suggère la fonction lagrangienne

$$L(x, \lambda) = f(x) + (\lambda, Bx - c) \qquad \forall x \in \mathbb{R}^n, \quad \forall \lambda \geq 0 \text{ et } \lambda \in \mathbb{R}^m.$$

ce qui transforme le problème primal en un problème de point-selle

$$\min_{Bx \leq c} \; f(x) = \min_{x \in \mathbb{R}^n} \max_{\lambda \geq 0} L(x, \lambda)$$

Contraintes de la forme: $Bx \leq c$ (suite)

La condition d'optimalité du lagrangien sera

- $(\frac{\partial L}{\partial x}, x - \bar{x}) \geq 0 \quad \forall x \in \mathbb{R}^n$ ce qui donne

$$\frac{\partial L}{\partial x} = 0 \Longleftrightarrow \nabla f(\bar{x}) + B^t \lambda = 0$$

- $(\frac{\partial L}{\partial \lambda}, \lambda - \bar{\lambda}) \leq 0 \quad \forall \lambda \geq 0$. Etant donné que $\{\lambda \geq 0\}$ forme un cône convexe, on obtient les conditions suivantes

$$(\frac{\partial L}{\partial \lambda}, \lambda) = (B\bar{x} - c, \lambda) \leq 0 \quad \forall \lambda \geq 0 \quad \Longrightarrow \quad B\bar{x} - c \leq 0 \Longleftrightarrow B\bar{x} \leq c$$
$$(\frac{\partial L}{\partial \lambda}, \bar{\lambda}) = 0 \Longleftrightarrow (B\bar{x} - c, \bar{\lambda}) = 0$$

Conditions KKT pour les contraintes: $Bx \leq c$

A partir des conditions d'optimalité du lagrangien $L(x, \lambda)$, on peut écrire les conditions dites de Karush-Kuhn-Tucker (KKT) pour le problème

$$\min_{Bx \leq c} \quad f(x)$$

Observons que $(B\bar{x} - c, \bar{\lambda}) = 0$ avec $\bar{\lambda} \geq 0$ signifie que $\bar{\lambda}_i \, (B\bar{x} - c)_i = 0$ pour tous les $i = 1, \ldots, m$.

Les conditions KKT s'écrivent sous la forme:

$$\left\{ \begin{array}{rcll} \nabla f(x) + B^t \lambda & = & 0, & \\ \lambda_i & \geq & 0, & \forall i = 1, \ldots, m, \\ (Bx)_i & \leq & c_i, & \forall i = 1, \ldots, m, \\ \lambda_i \, (Bx - c)_i, & = & 0 & \forall i = 1, \ldots, m. \end{array} \right.$$

Point de vue des cônes duaux: $Bx \leq c$

Le problème $f(a) = \min\limits_{Bx \leq c} \ f(x)$ est équivalent à

$$\min_{-Bx \geq -c} \ f(x)$$

De nouveau, on fait l'hypothèse que $c \in Im\, B$ donc il existe x_0 tel que $Bx_0 = c$. Ceci implique

$$\begin{aligned} U &= \{x \mid Bx \leq c\} = \{x \mid -Bx \geq -c\} \\ &= \{x \mid -B(x - x_0) \geq 0\} = x_0 + \{-Bx \geq 0\} = x_0 + W \end{aligned}$$

où $W = \{-Bx \geq 0\}$ est le cône positif.
La condition d'optimalité du problème s'écrit

$$(\nabla f(a), x - a) \geq 0 \qquad \forall x \in U \quad \Longrightarrow (\nabla f(a), y) \geq 0 \quad \forall y \in W,$$

autrement dit $\nabla f(a) \in W^*$. Mais $W^* = -B^t \Lambda^+$ avec $\Lambda^+ = \{\lambda \geq 0\}$.
Ceci montre l'existence du multiplicateur λ vérifiant $\nabla f(a) + B^t \lambda = 0$.

Point de vue des cônes duaux: (suite)

Mais il y a une autre condition que l'on peut tirer de la condition $(\nabla f(a), x - a) \geq 0$. Il s'agit de la condition

$$(\nabla f(a), a - x_0) = 0 \Longrightarrow (-B^t \lambda, a - x_0) = 0 \Longrightarrow (\lambda, Ba - Bx_0) = (\lambda, Ba - c) = 0.$$

En effet, il suffit de prendre $x = x_0$ et $x = x_0 + 2(a - x_0)$ dans le condition ci-dessus.

En combinant toutes les conditions ci-dessus, on obtient exactement le même système KKT que doit vérifier la solution a et le multiplicateur λ

$$\left\{ \begin{array}{rcll} \nabla f(x) + B^t \lambda & = & 0, & \\ \lambda_i & \geq & 0, & \forall i = 1, \dots, m, \\ (Bx)_i & \leq & c_i, & \forall i = 1, \dots, m, \\ \lambda_i \, (Bx - c)_i & = & 0, & \forall i = 1, \dots, m. \end{array} \right.$$

Cas général: contraintes $g(x) = 0$

Appliquons la méthode de Lagrange au problème

$$\min_{g(x)=0} \quad f(x)$$

où $g = (g_1, g_2, \dots, g_m)$ désigne m contraintes scalaires. En général, l'ensemble des contraintes n'est pas convexe.

On prend la fonction lagrangienne

$$L(x, \lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m.$$

Ceci transforme le problème primal en un problème de point-selle

$$\min_{g(x)=0} \quad f(x) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} L(x, \lambda)$$

La condition d'optimalité du lagrangien sera

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \Longleftrightarrow \nabla f(x) + \displaystyle\sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0, \\ \dfrac{\partial L}{\partial \lambda} = 0 \Longleftrightarrow g_i(x) = 0, \qquad \forall i = 1, \dots, m. \end{cases}$$

Point de vue des cônes duaux: cas général

Cône des directions admissibles $C_U(x)$**:**
Soit $U \subset \mathbb{R}^n$ un sous-ensemble fermé. Fixons un point $x \in U$. On dira que $h \in C_U(x)$ est direction admissible s'il existe une courbe $x(t) \in U$ définie localement pour des valeurs positives de $t \geq 0$ et passant par le point $x = x(0)$ de sorte que

$$\frac{dx}{dt}(0) = \lim_{\substack{t \to 0 \\ t > 0}} \frac{x(t) - x}{t} = h$$

Propriétés de $C_U(x)$**:**

- $C_U(x)$ est un cône fermé de sommet 0 (pas nécessairement convexe)
- Si U est convexe, $C_U(x)$ est un cône convexe et

$$C_U(x) = \text{ la fermeture de } \{\lambda(y - x) \mid y \in U, \lambda \geq 0\}$$

Condition d'optimalité du premier ordre: cas général

Théorème: Condition nécessaire d'optimalité du premier ordre

Soit $f : U \longrightarrow \mathbb{R}$ une fonction numérique de classe C^1 définie sur un ensemble fermé $U \subset \mathbb{R}^n$. Si $a \in U$ est un point minimisant de f,

$$f(a) = \min_{x \in U} f(x),$$

alors

$$\forall h \in C_U(a) \quad \Longrightarrow (\nabla f(a), h) \geq 0 \qquad \Longleftrightarrow \nabla f(a) \in C_U^*(a)$$

Preuve: $h \in C_U(a) \Longrightarrow \exists x(t) \in U$ tel que $x(0) = a, \quad \frac{dx}{dt}(0) = h.$
On a que

$$\begin{aligned}
f(x(t)) - f(a) &\geq 0 \\
\Longrightarrow \quad \frac{f(x(t)) - f(x(0))}{t} &\geq 0 \qquad \forall t > 0 \\
\Longrightarrow \quad \lim_{t \to 0} \frac{f(x(t)) - f(x(0))}{t} &\geq 0 \\
(\nabla f(a), \tfrac{dx}{dt}(0)) &\geq 0 \Longrightarrow (\nabla f(a), h) \geq 0 \Longleftrightarrow \nabla f(a) \in C_U^*(a)
\end{aligned}$$

Contraintes d'égalité

Soient $g_1, g_2, \ldots, g_m$ m fonctions numériques de classe C^1. On pose

$$U = \{x \in \mathbb{R}^n \mid g_1(x) = g_2(x) = \cdots = g_m(x) = 0\}$$

En général U n'est pas convexe mais toujours fermé. Il est commode d'introduire l'application

$$\begin{array}{rcl} g : \mathbb{R}^n & \rightarrow & \mathbb{R}^m \\ x & \rightarrow & (g_1(x), g_2(x), \ldots, g_m(x)) \end{array}$$

Définition: La matrice jacobienne au point x est définie par

$$(Dg(x))_{ij} = \frac{\partial g_i}{\partial x_j}(x)$$

qui est de format $m \times n$.

Définition: Un point x est dit régulier si $rg(Dg(x)) = m$.

Théorème: Si x est un point régulier de g, alors le cône des directions admissibles est donné par

$$C_U(x) = Ker\ Dg(x)$$

Contraintes d'égalité (suite)

Appliquons ce résultat au problème

$$f(a) = \min_{g(x)=0} \quad f(x)$$

où $g = (g_1, g_2, \dots, g_m)$ désigne m contraintes scalaires.

Condition nécessaire d'optimalité du premier ordre est

$$\forall h \in C_U(a) \quad (\nabla f(a), h) \geq 0 \qquad \Longleftrightarrow \qquad \nabla f(a) \in C_U^*(a)$$

Or $C_U^*(a) = (Ker\, Dg(a))^* = (Ker\, Dg(x))^\perp = Im\,(Dg(a))^t$. On obtient ainsi

$$\nabla f(a) \in Im\,(Dg(a))^t$$

Autrement dit, il existe $\lambda \in \mathbb{R}^m$ tel que

$$\nabla f(a) + \sum_{i=1}^{m} \lambda_i \nabla g_i(a) = 0.$$

Contraintes d'égalité (suite)

En résumé, le point $a \in U$ doit vérifier le système non linéaire

$$\begin{cases} f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0, & \\ g_i(x) = 0, & \forall i = 1, \dots, m. \end{cases}$$

qui correspond à la condition d'optimalité du lagrangien

$$L(x, \lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m.$$

Contraintes de la forme: $g(x) \leq 0$

Appliquons la dualité lagrangienne au problème (primal)

$$\min_{g(x) \leq 0} \quad f(x)$$

où $g(x) = (g_1(x), g_2(x), \ldots, g_m(x))$.

Lemme: La fonction indicatrice de $K = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ est donnée par

$$I_K(x) = \max_{\lambda \geq 0} (\lambda, g(x))$$

Démonstration: en effet, si $g_i(x) > 0$, on peut prendre $\lambda = 0$ sauf pour la composante i

$$\lambda_i = r\,(g_i(x)) > 0 \Longrightarrow (\lambda, g(x)) = r\, g_i(x)^2 \to \infty \text{ si } r \to \infty.$$

Le maximum sera ∞.
Au contraire, si $g(x) \leq 0$, on a que $(\lambda, g(x)) \leq 0 \quad \forall \lambda \geq 0$ car $g(x) \leq 0$. Donc le maximum sera 0.

Contraintes de la forme: $g(x) \leq 0$ (suite)

Par conséquent, on peut écrire

$$\min_{g(x)\leq 0} \; f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \geq 0}(\lambda, g(x))$$

Ceci suggère la fonction lagrangienne

$$L(x,\lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \geq 0 \text{ et } \lambda \in \mathbb{R}^m.$$

ce qui transforme le problème primal en un problème de point-selle

$$\min_{g(x)\leq 0} \; f(x) = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0} L(x,\lambda)$$

Contraintes de la forme: $g(x) \leq 0$ (suite)

La condition d'optimalité du lagrangien sera

- $(\frac{\partial L}{\partial x}, x - \bar{x}) \geq 0 \quad \forall x \in \mathbb{R}^n$ ce qui donne

$$\frac{\partial L}{\partial x} = 0 \Longleftrightarrow \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0$$

- $(\frac{\partial L}{\partial \lambda}, \lambda - \bar{\lambda}) \leq 0 \quad \forall \lambda \geq 0$. Etant donné que $\{\lambda \geq 0\}$ forme un cône convexe, on obtient les conditions suivantes

$$(\frac{\partial L}{\partial \lambda}, \lambda) = (g(\bar{x}), \lambda) \leq 0 \quad \forall \lambda \geq 0 \quad \Longrightarrow \quad g(\bar{x}) \leq 0$$
$$(\frac{\partial L}{\partial \lambda}, \bar{\lambda}) = 0 \Longleftrightarrow (g(\bar{x}), \bar{\lambda}) = 0$$

Conditions KKT pour les contraintes: $g(x) \leq 0$

A partir des conditions d'optimalité du lagrangien $L(x, \lambda)$, on peut écrire les conditions dites de Karush-Kuhn-Tucker (KKT) pour le problème

$$\min_{g(x) \leq 0} \quad f(x)$$

Observons que $(g(\bar{x}), \bar{\lambda}) = 0$ avec $\bar{\lambda} \geq 0$ signifie que $\bar{\lambda}_i \, g_i(\bar{x}) = 0$ pour tous les $i = 1, \ldots, m$.

Les conditions KKT s'écrivent sous la forme:

$$\left\{ \begin{array}{rcll} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) & = & 0, & \\ \lambda_i & \geq & 0, & \forall i = 1, \ldots, m, \\ g_i(x) & \leq & 0, & \forall i = 1, \ldots, m, \\ \lambda_i \, g_i(x) & = & 0 & \forall i = 1, \ldots, m. \end{array} \right.$$

Point de vue des cônes duaux: $g(x) \leq 0$

On désire traiter du problème

$$f(a) = \min_{g(x) \leq 0} \; f(x)$$

du point de vue des cônes duaux. On pose $U = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$. En général, U n'est pas convexe. Par contre, si les g_i sont convexes, l'ensemble U est convexe.
Nous savons que l'optimum $a \in U$ est caractérisé par la condition d'optimalité

$$(\nabla f(a), h) \geq 0 \qquad \forall h \in C_U(a)$$

où $C_U(a)$ est le cône des directions admissibles au point a.

Définition: l'ensemble des contraintes actives au point x_0 est noté par

$$I(x_0) = \{j = 1, 2, \ldots, m \mid g_j(x_0) = 0\}$$

Son complémentaire forme l'ensemble des contraintes inactives.

Point de vue des cônes duaux: $g(x) \leq 0$ (suite)

L'objectif est le calcul du cône des directions admissibles $C_U(x)$ pour $U = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$.

Théorème:

$$C_U(x_0) \subset \{h \in \mathbb{R}^n \mid (\nabla g_j(x_0), h) \leq 0 \quad \forall j \in I(x_0)\}.$$

Démonstration:
Soit $h \in C_U(x_0) \Longrightarrow \exists x(t) \in U$ tel que $x(0) = x_0, \quad \frac{dx}{dt}(0) = h.$
Pour $j \in I(x_0)$, on a

$$\begin{array}{rcll}
g_j(x(t)) & \leq & 0 & \\
\Longrightarrow g_j(x(t)) - g_j(x_0) & \leq & 0 & \text{car } g_j(x_0) = 0 \\
\Longrightarrow \dfrac{g_j(x(t)) - g_j(x(0))}{t} & \leq & 0 & \forall t > 0 \\
\Longrightarrow \lim\limits_{t \to 0} \dfrac{g_j(x(t) - g_j(x(0))}{t} & \leq & 0 & \\
(\nabla g_j(x_0), \frac{dx}{dt}(0)) & \leq & 0 & \Longrightarrow (\nabla g_j(x_0), h) \leq 0
\end{array}$$

Point de vue des cônes duaux: $g(x) \leq 0$ (suite)

Définition: Le point x_0 vérifie l'hypothèse de qualification si on a égalité

$$C_U(x_0) = \{h \in \mathbb{R}^n \mid (\nabla g_j(x_0), h) \leq 0 \quad \forall j \in I(x_0)\}.$$

Dans la pratique, on utilisera le critère que la matrice B ci-dessous doit être de rang maximal.

Ceci est équivalent à

$$C_U(x_0) = \{h \in \mathbb{R}^n \mid Bh \leq 0\}$$

où

$$B = \begin{pmatrix} \nabla g_{j_1}(x_0) \\ \nabla g_{j_2}(x_0) \\ \vdots \\ \nabla g_{j_p}(x_0) \end{pmatrix}_{p \times n}$$

avec $I(x_0) = \{j_1, j_2, \ldots, j_p\}$.

Point de vue des cônes duaux: $g(x) \leq 0$ (suite)

La condition d'optimalité du problème

$$f(a) = \min_{g(x) \leq 0} \quad f(x)$$

s'écrit

$$\nabla f(a) \in C_U^*(a) = -B^t \Lambda^+ = \{-B^t \lambda \mid \lambda \geq 0\}.$$

Autrement dit, il existe $\lambda \in \mathbb{R}^m$ avec $\lambda \geq 0$ tel que

$$\left\{ \begin{array}{rcll} \nabla f(a) + \sum_{i=1}^{m} \lambda_i \nabla g_i(a) & = & 0, & \\ \lambda_i & \geq & 0, & \forall i = 1, \dots, m, \\ g_i(a) & \leq & 0, & \forall i = 1, \dots, m, \end{array} \right.$$

Il reste à montrer que $\lambda_i \, g_i(a) = 0$. En effet, si $i \in I(x_0) \Longrightarrow g_i(a) = 0$. Sinon, $i \notin I(x_0)$ on pose $\lambda_i = 0$ car ∇g_i n'apparait pas dans la matrice B.

Méthodes numériques: contraintes d'égalité

Considérons le problème de minimisation ayant seulement des contraintes d'égalité

$$\min_{g(x)=0} \quad f(x)$$

où $g = (g_1, g_2, \ldots, g_m)$ désigne m contraintes scalaires.

En gros, il y a deux classes de méthodes numériques pour ce type de problème.

- Approche couplée: basée sur la méthode des multiplicateurs de Lagrange, résolution en (x, λ).
- Approche découplée: basée sur la théorie de la dualité, résolution en x seulement.

Méthodes numériques: contraintes d'égalité

L'idée consiste à résoudre directement le système d'équations obtenu par la méthode des multiplicateurs de Lagrange. Si on introduit la fonction lagrangienne

$$L(x,\lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m.$$

la condition d'optimalité du problème $\min\limits_{g(x)=0} \; f(x)$ s'écrit

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0, \\ g_i(x) = 0, \qquad \forall i = 1, \dots, m. \end{cases}$$

Ces équations forment un système de $n+m$ équations par rapport aux $n+m$ variables x et λ.
Notons ce système par $F(x,\lambda) = (F_x(x,\lambda), F_\lambda(x,\lambda)) = (0,0)$ où

$$F_x(x,\lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x),$$
$$F_\lambda(x,\lambda) = g(x) = (g_1(x), g_2(x), \dots, g_m(x))^t.$$

Méthodes numériques: contraintes d'égalité (suite)

Le système

$$F(x,\lambda)=0 \Longleftrightarrow \left\{ \begin{array}{l} F_x(x,\lambda)=0 \\ F_\lambda(x,\lambda)=0 \end{array} \right.$$

sera résolu par l'algorithme de Newton

1. Résoudre: $DF(X_k)\,\Delta X=-F(X_k)$,
2. $X_{k+1}=X_k+\Delta X$,

où $X=(x,\lambda)$ et $DF(X)$ est la matrice tangente (jacobienne) au point X du système.

Calculons la matrice tangente qui a la forme d'une matrice 2×2 par blocs

$$DF(X)=\left[\begin{array}{cc} \dfrac{\partial F_x}{\partial x} & \dfrac{\partial F_x}{\partial \lambda} \\ \dfrac{\partial F_\lambda}{\partial x} & \dfrac{\partial F_\lambda}{\partial \lambda} \end{array}\right]$$

On a les formules

- $\dfrac{\partial F_x}{\partial x}=H_f+\sum_{i=1}^{m}\lambda_i H_{g_i}$,
- $\dfrac{\partial F_\lambda}{\partial x}=Dg$ où Dg est la matrice jacobienne $m\times n$ de l'application g,
- $\dfrac{\partial F_x}{\partial \lambda}=Dg^t$ et $\dfrac{\partial F_\lambda}{\partial \lambda}=0$.

Méthode de pénalisation

En utilisant la fonction indicatrice de l'ensemble $U = \{g(x) = 0\}$, nous pouvons enlever les contraintes d'égalité

$$\min_{g(x)=0} f(x) = \min_{x} f(x) + I_U(x)$$

Le principe de la méthode de pénalisation consiste à approcher la fonction indicatrice par une fonction plus régulière $I_U(x) \approx r\,\Psi_r(x)$ où r est un paramètre qui tend vers ∞. La fonction $\Psi_r(x)$ vérifie les propriétés

1. $\Psi_r(x) \geq 0 \qquad \forall x \in \mathbb{R}^n \qquad \forall r > 0,$
2. $\Psi_r(x) = 0 \quad \Longleftrightarrow \quad x \in U \qquad \forall r > 0,$
3. $\lim_{r \to \infty} \Psi_r(x) = 0 \quad \Longleftrightarrow \quad x \notin U.$

Méthode de pénalisation (suite)

Pour les contraintes d'égalité, la fonction $\Psi_r(x)$ la plus utilisée est

$$\Psi_r(x) = \frac{r}{2} \|g(x)\|^2$$

Par conséquent, le problème pénalisé s'écrit

$$\min_{x \in \mathbb{R}^n} \tilde{f}_r(x) = \min_{x \in \mathbb{R}^n} f(x) + \frac{r}{2} ||g(x)||^2$$

où $r > 0$ est un paramètre réel, en général grand, fixé d'avance.

On calcule le point minimum grâce à la condition d'optimalité

$$\nabla \tilde{f}_r(x) = 0.$$